

**Multilevel algorithm for flow observables in gauge theories**Miguel García Vera<sup>1,2</sup> and Stefan Schaefer<sup>1</sup><sup>1</sup>*John von Neumann Institute for Computing (NIC), DESY, Platanenallee 6, D-15738 Zeuthen, Germany*<sup>2</sup>*Institut für Physik, Humboldt Universität zu Berlin, Newtonstraße 15, D-12489 Berlin, Germany*

(Received 29 January 2016; published 5 April 2016)

We study the possibility of using multilevel algorithms for the computation of correlation functions of gradient flow observables. For each point in the correlation function, an approximate flow is defined which depends only on links in a subset of the lattice. Together with a local action, this allows for independent updates and consequently a convergence of the Monte Carlo process faster than the inverse square root of the number of measurements. We demonstrate the feasibility of this idea in the correlation functions of the topological charge and the energy density.

DOI: 10.1103/PhysRevD.93.074502

**I. INTRODUCTION**

In Monte Carlo simulations of Yang-Mills gauge theories, correlation functions of gluonic operators suffer from a severe signal-to-noise problem [1]. While the signal of a two-point function itself falls off exponentially with the distance between the operators, the variance is largely independent of their separation. Since the error decreases only like  $1/\sqrt{N}$ , with  $N$  the number of measurements, this makes their measurement in numerical calculations for large separations exceedingly difficult.

Due to their favorable renormalization properties, correlation functions of observables defined through the Yang-Mills gradient flow are an important tool to study gauge theories [2–4]. In particular, they allow for a computationally economical definition of the topological susceptibility on the lattice,

$$\chi_{\text{top}} = \frac{1}{V} \int dx dy \langle q(x)q(y) \rangle,$$

with the topological charge density  $q(x)$  defined through the Wilson flow.

The signal-to-noise problem present at large distances in the  $\langle q(x)q(y) \rangle$  correlation function translates into a lack of volume averaging of  $\chi_{\text{top}}$ : the statistical error of the susceptibility from a given number of configurations does not improve with increasing volume. In pure gauge theory, this can be partially overcome with large statistics, but in practice, rather small lattices are still used, and the finite size effect needs to be carefully controlled [5]. In large volume, it is therefore beneficial to study the dependence of  $\langle q(x)q(y) \rangle$  on  $|x - y|$  directly and model the large distance behavior [6] or integrate it such that the contribution of the tail can be neglected [7] given the statistical accuracy.

On a related note, we also point out that it has been suggested to extract the masses of glueballs from the large distance behavior of two-point functions of the smoothed topological charge and energy density [8]. For this

approach to work, it is highly beneficial to have a precise determination of the tail of the correlator at large distances.

One way to deal with the signal-to-noise problem is to use multilevel algorithms which rely on the locality of the observable and of the action [9,10]. If it is possible to decompose the observables in contributions from different parts of the lattice, each of them can be updated independently. Depending on the number of sublattices and the efficiency of the decomposition, the signal-to-noise problem can be eliminated or at least reduced substantially. Recently, such a type of algorithms has also been adapted to the case of quenched lattice QCD to compute fermionic correlators [11].

In the case of flow observables, multilevel algorithms cannot be applied directly basically due to the fact that the flow has a footprint which is not finite. In this paper, we propose a first step in the direction of solving this problem. We study a two-level algorithm where the lattice is decomposed into two subvolumes and observables are defined such that they depend only on the fields in the respective subvolume.

The paper is organized as follows. In Sec. II, we describe the algorithm, and in Sec. III, we define the observables that we use in this study. Then, in Sec. IV, we demonstrate the feasibility of this setup and discuss how the improvement works before we conclude.

**II. ALGORITHM**

In order to make the discussion of the algorithm as self-contained as possible, we shall briefly present the main ideas introduced in Refs. [10,12], in a context which is directly applicable to our case.

**A. Factorized observables**

For simplicity, we consider  $SU(N)$  Yang-Mills gauge theory on the lattice with the standard Wilson action, although a more general type of actions can be used,

$$S[U] = \frac{\beta}{N} \sum_P \text{Tr}\{1 - U(P)\}, \quad (1)$$

where  $U(P)$  is the product of gauge links around the plaquette  $P$ .

Take  $B$ ,  $L$ , and  $R$  to be three disjoint subsets of gauge links, such that they make up for the totality of gauge links on the lattice. We choose  $B$  in such a way that the gauge action  $S[U_L, U_B, U_R]$  can be decomposed as  $S_L[U_L, U_B] + S_R[U_B, U_R] + S_B[U_B]$ , where by  $U_{L,B,R}$  we refer to the set of gauge links which belong to  $L$ ,  $B$ , and  $R$  respectively. One natural choice for  $B$  is the subset of all spatial links at a fixed time slice  $x_0^B$ , so that  $L$  and  $R$  are simply defined as all gauge links that are located to the left or to the right of the boundary  $B$ . This setup is depicted in Fig. 1.

For two observables  $\mathcal{O}(x)$  and  $\mathcal{O}'(y)$ , which are defined for  $x \in L$  and  $y \in R$ , the decomposition of the action makes it possible to write

$$\begin{aligned} \langle \mathcal{O}\mathcal{O}' \rangle &= \int dU_B p(B) [\mathcal{O}]_L [\mathcal{O}']_R, \\ [\mathcal{A}]_{L,R} &= \frac{1}{Z_{L,R}} \int dU_{L,R} \mathcal{A} e^{-S_{L,R}[U_{L,R}, U_B]}, \end{aligned} \quad (2)$$

where  $\mathcal{A}$  is either  $\mathcal{O}$  or  $\mathcal{O}'$ ,  $Z_L$  and  $Z_R$  are the normalization factors such that  $[1]_{L,R} = 1$ , and  $p(B) = \frac{Z_L Z_R}{Z} e^{-S[U_B]}$ , with  $Z$  the standard partition function.

Equation (2) expresses the fact that one can average an observable over  $L$  and  $R$  independently while keeping  $B$  fixed and then take the average over the possible values of  $B$ . As discussed in Refs. [10,12], this process can be iterated if the operators  $\mathcal{O}$  or  $\mathcal{O}'$  can be subsequently factorized. This is the property of factorization that was exploited originally in Ref. [10] to show an exponential reduction in the error of the expectation value of large Wilson loops.

The idea presented above can be realized in a Monte Carlo simulation as follows. First, generate  $N_0$  regular updates which are used to perform the integration over  $U_B$  in Eq. (2). Then, for each of the  $N_0$  original configurations,  $N_1$  updates of  $L$  and  $R$  are done independently while keeping  $B$  fixed, so that for the product  $[\mathcal{O}]_L [\mathcal{O}']_R$ , the error decreases ideally as  $1/N_1$  instead of the standard  $1/\sqrt{N_1}$ . As shown in the Appendix, this can be reached only for operators with vanishing expectation value  $\langle \mathcal{O} \rangle = \langle \mathcal{O}' \rangle = 0$ . Therefore, in the following, we restrict ourselves to the connected correlation functions.

Note that factorization makes it possible to obtain a better scaling for the errors in  $[\mathcal{O}]_L [\mathcal{O}']_R$ , but the error on

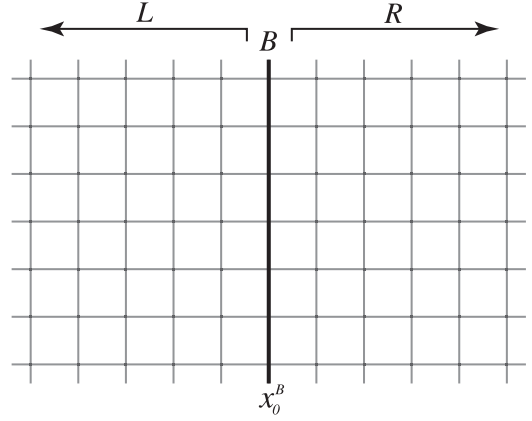


FIG. 1. Factorized lattice setup. The lattice is split into two subvolumes  $L$  and  $R$  which are separated by the boundary links  $B$  defined as the spatial links at the time slice  $x_0^B$ .

the final expectation value  $\langle \mathcal{O}\mathcal{O}' \rangle$  depends on the average over  $B$  which scales as  $1/\sqrt{N_0}$ . This means that for large values of  $N_1$ , the error is controlled by the fluctuations of  $B$ , and hence the dominant scaling will be the  $1/\sqrt{N_0}$ . As discussed in the following sections and as shown in the Appendix, in practice one can take very large values of  $N_1$  before the ideal scaling is no longer valid.

## B. Modified flow observables

Given the gauge link variables  $U(x, \mu)$ , the flow variables  $V^t(x, \mu)$  associated to them are defined by the equation [2,3,13]

$$\begin{aligned} \dot{V}^t(x, \mu) &= -g_0^2 \{ \partial_{x, \mu} S(V) \} V^t(x, \mu), \\ V^{t=0}(x, \mu) &= U(x, \mu). \end{aligned} \quad (3)$$

The effect of the flow can be viewed as a smoothing of the gauge fields over a spherical range with a mean square radius of  $\sqrt{8t}$ . Because of this, any observable defined in  $L$  or  $R$  has a nontrivial dependence on gauge links from the opposite domain at positive flow time  $t$ , and it cannot be factorized as required for Eq. (2) to hold. However, the smoothing produced by the flow is exponentially suppressed at large distances, which leads us to propose a slightly modified version of the flow equations, such that an observable computed with the modified flow gauge links  $\tilde{V}^t$  is a good approximation to the original one and can be factorized as required in Eq. (2).

If the Wilson action is also used in the definition of the flow, we propose the following modified flow equation:

$$\dot{\tilde{V}}^t(x, \mu) = \begin{cases} -g_0^2 \{ \partial_{x, \mu} S(\tilde{V}^t) \} \tilde{V}^t(x, \mu), & \tilde{V}^{t=0}(x, \mu) = U(x, \mu), \quad \text{if } U(x, \mu) \in L \cup R. \\ U(x, \mu), & \text{if } U(x, \mu) \in B. \end{cases} \quad (4)$$

The modified version accounts for integrating the flow equations while the links at  $B$  are kept fixed. It is constructed such that for each link  $U(x, \mu) \in L$ , also the smoothed link  $\tilde{V}^t(x, \mu)$  only depends on links in  $L$  and  $B$ . Therefore, for an observable  $\mathcal{O}(x)$ , in either  $L$  or  $R$ , the modified flow observable  $\tilde{\mathcal{O}}^t(x)$  does not get any contribution from the links in the opposite domain. If  $\tilde{\mathcal{O}}^t(x)$  is a good approximation of  $\mathcal{O}^t(x)$ , one can take advantage of factorization to obtain a better scaling of the errors of  $\langle \mathcal{O} \mathcal{O}' \rangle$  with respect to the  $N_1$  nested Monte Carlo updates.

### C. Two-point correlation function

We now consider the case of the connected two-point correlation function  $\mathcal{O}(x)\mathcal{O}(y)$  for  $x$  and  $y$  spacetime points in the four dimensional lattice and put together the modified flow observables with the multilevel scheme. We define

$$\begin{aligned} C_{\mathcal{O}}^t(x, y) &= \langle \mathcal{O}^t(x)\mathcal{O}^t(y) \rangle_C \\ &= \langle \mathcal{O}^t(x)\mathcal{O}^t(y) \rangle - \langle \mathcal{O}^t(x) \rangle \langle \mathcal{O}^t(y) \rangle \end{aligned} \quad (5)$$

as the connected ( $C$ ) correlation function of the observable  $\mathcal{O}$ . If the two points  $x$  and  $y$  are separated from the boundary  $B$  by a distance much larger than the radius of the flow  $\sqrt{8t}$ , then the modified version of Eq. (5) using the gauge links  $\tilde{V}^t$  is a good approximation to the original correlator. To show this, we look at the correction term  $\Delta$ , defined as the difference between the flow observable and the observable computed using the modified flow

$$\Delta_{\mathcal{O}}^t(x, y) = C_{\mathcal{O}}^t(x, y) - \tilde{C}_{\mathcal{O}}^t(x, y). \quad (6)$$

Notice that we have left the dependence on both  $x$  and  $y$  explicit, as the presence of the boundary  $B$  breaks full translation invariance and one must keep track not only of the distance  $|x - y|$  between source and sink but also of the distance of both  $x$  and  $y$  with respect to  $B$ . The reason for this will become evident in the next section when we discuss a practical application of the algorithm. When not explicitly needed, we will drop the  $t$  index in every quantity.

For the observables discussed in the next section, our data show that for a sufficiently large separation from  $B$  compared to the smoothing radius,  $\Delta$  becomes negligible. In spite of that, our strategy is not to neglect the correction term  $\Delta$ . Instead, in a nested Monte Carlo simulation, the idea is to use first the  $N_0$  generated configurations to estimate  $\Delta$  and then use this estimation to correct for the value of  $\tilde{C}_{\mathcal{O}}(x, y)$ . For this to work, we need that the fluctuations of  $\Delta$  are much smaller than the fluctuations of  $C_{\mathcal{O}}$  in such a way that we can use the  $N_0$  updates to estimate  $\Delta$  and subsequently perform the  $N_1$  nested Monte Carlo updates independently in  $L$  and  $R$  to compute  $\tilde{C}_{\mathcal{O}}$ .

The main equation of this paper is a modified version of Eq. (2), which takes into account the correction term  $\Delta$  and is applicable for any two-point correlation function of Wilson flow observables. We define an estimator  $\hat{C}_{\mathcal{O}}^t(x, y)$  of  $C_{\mathcal{O}}^t(x, y)$  as

$$\begin{aligned} \hat{C}_{\mathcal{O}}^t(x, y) &= \frac{1}{N_0} \sum_{N_0} \{ [\tilde{\mathcal{O}}^t(x)]_L [\tilde{\mathcal{O}}^t(y)]_R + \Delta_{\mathcal{O}}^t(x, y) \} \\ [\tilde{\mathcal{O}}^t(z)]_{L,R} &= \frac{1}{N_1} \sum_{N_1} \tilde{\mathcal{O}}^t(z), \quad z = x, y, \end{aligned} \quad (7)$$

where  $(x, y) \in L \times R$ . The estimator in Eq. (7) is correct up to errors of order  $O(1/\sqrt{N_0})$ , which comes from the fact that  $\Delta$  is only computed on the  $N_0$  standard updates. However, in the next section, we show that the fluctuations of  $\Delta$  are exponentially suppressed with the distance to the boundary, so that the leading term for the scaling of the error in  $\hat{C}$  comes from the correlator of the modified flow observables.

### III. NUMERICAL TEST OF THE MODIFIED FLOW OBSERVABLES

To test our algorithm, we work with the SU(3) gauge group and a set of gauge configurations generated with the parameters shown in Table I. The configurations are generated for a value of  $\beta = 6.11$ , which corresponds to a lattice spacing of  $a \approx 0.08$  fm and an effective smearing radius  $\sqrt{8t_0} \approx 6a$ . Open boundary conditions are used in the time direction [14]. We consider two observables, the topological charge density  $q$  and the Yang-Mills energy density  $e$ . In particular, we look at the connected two-point correlation function of the time slice summed  $\bar{q}$  and  $\bar{e}$ ,

$$\begin{aligned} C_q^t(x_0, r) &= \langle \bar{q}^t(x_0)\bar{q}^t(x_0+r) \rangle_C, \quad \bar{q}^t(x_0) = a^3 \sum_{\vec{x}} q^t(\vec{x}, x_0) \\ C_e^t(x_0, r) &= \langle \bar{e}^t(x_0)\bar{e}^t(x_0+r) \rangle_C, \quad \bar{e}^t(x_0) = a^3 \sum_{\vec{x}} e^t(\vec{x}, x_0), \end{aligned} \quad (8)$$

where we have left the  $x_0$  dependence explicit in order to keep track of the distance to the boundary  $B$ , which is chosen to be the subset of spatial links with time coordinate  $x_0^B = T/2$ . All computations are done in such a way that

TABLE I. Lattice parameters. We report the lattice bare coupling  $\beta$ , the lattice dimensions  $L$  and  $T$ , the scale parameter  $t_0$  defined in Ref. [3], the lattice spacing  $a$  computed using the  $r_0 = 0.5$  [fm] scale from [15], and the number of generated configurations  $N_0$ .

$\beta$	$(T/a) \times (L/a)^3$	$t_0/a^2$	$a$ (fm)	$N_0$
6.11	$80 \times 20^3$	4.5776(15)	0.078	384

both  $x_0$  and  $x_0 + r$  are placed far enough from the open boundaries. From now on, we shall use  $\mathcal{O}$  to refer to either  $q$  or  $e$  when there is no need to make a distinction between them.

We use the 384 independent configurations to study the dependence of the fluctuations of  $\Delta$  on both  $x_0$  and  $r$ . First, we consider the correlators which are symmetric with respect to  $B$ , so we choose a source which is placed at the value of  $x_0 = (T - r)/2$ . In this case, the correlator is only a function of  $r$  and given by  $C_{\mathcal{O}}(r) = C_{\mathcal{O}}((T - r)/2, r)$ .

Figure 2 shows the dependence of the error of both  $C_{\mathcal{O}}(r)$  and  $\Delta_{\mathcal{O}}(r)$  for a fixed value of the flow time  $t = t_0$ . The errors are computed by measuring the autocorrelation function as described in Ref. [16]. We note that for separations from the boundary larger than the smoothing radius, the fluctuations in  $\Delta$  are below 5% of those of the observable. As will be discussed in Sec. IV B, the fact that the ratio between the fluctuations of the observable and those of the correction term decrease at large distances contributes to the fact that the algorithm is efficient up to very large values of  $N_1$ .

Since the effective smearing radius produced by the flow grows as  $\sqrt{t}$ , the effect of freezing the boundary links at  $B$  increases monotonically with the flow time. We have observed this behavior in our data, but we are more interested in the behavior of the correlation functions at the reference flow scale  $t = t_0$ . For different values of the flow, a similar analysis can be performed. However, it is clear that if the fluctuations of  $\Delta$  are “small” for a given value of  $t'$ , they are also small for  $t < t'$ .

Next, we go beyond the symmetric case and look at the  $x_0$  dependence of  $\Delta$ . Figure 3 shows a plot of the

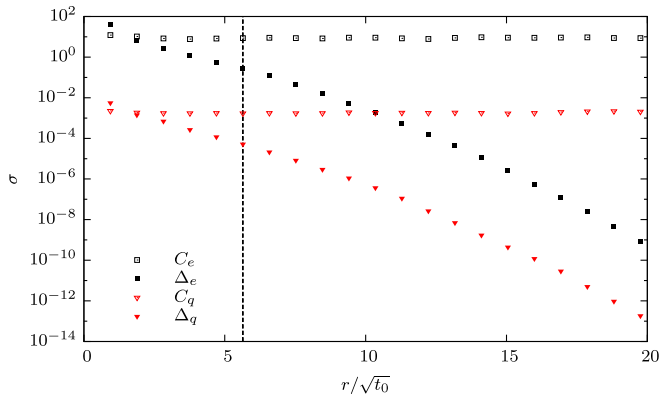


FIG. 2. Statistical error  $\sigma$  of  $\Delta_{\mathcal{O}}$  and  $C_{\mathcal{O}}$  at flow time  $t = t_0$ . For both observables,  $\bar{e}$  and  $\bar{q}$ , the error in the correlator is independent of the distance  $r$ , but the errors of  $\Delta$  seem to decay at least exponentially with the distance. The dotted vertical line is added as a reference to indicate the point where the distance from the boundary is equal to the smoothing radius  $\sqrt{8t_0}$ . For larger separations, the errors in  $\Delta_{\mathcal{O}}$  are below 5% those of  $C_{\mathcal{O}}$ . Uncertainties are smaller than the data markers.

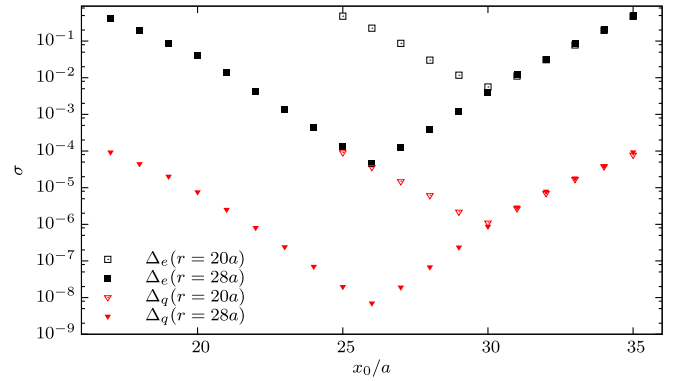


FIG. 3. Statistical error  $\sigma$  of  $\Delta_{\mathcal{O}}$  as a function of  $x_0$  for two values of  $r$  at flow time  $t = t_0$ . Open symbols correspond to a value of  $r = 20a = 9.4\sqrt{t_0}$ , while filled symbols correspond to a value of  $r = 28a = 13.2\sqrt{t_0}$ . The smallest error corresponds to the symmetric point in which both source and sink are placed far from the boundary. Uncertainties are smaller than the data markers.

errors in  $\Delta$  as a function of  $x_0$  for two fixed values of  $r$  at  $t = t_0$ .

Notice that the effect of the flow is that of a Gaussian smearing, so we should expect that the errors in  $\Delta$  decay at least exponentially with the distance to the boundary  $B$ . Both Figs. 2 and 3 show a behavior which is compatible with this statement.

The results presented in this section show that using the modified flow equations has little impact in the two-point function, and the effect can be incorporated in the correction term  $\Delta$ . When using Eq. (7), it is important to tune the value of  $N_0$  and  $N_1$  in such a way that the effect of  $\Delta$  remains under control. In particular, due to the exponential smoothing of the flow,  $N_1$  can be chosen larger at larger values of  $r$ , which is precisely where a higher precision is required.

## IV. RESULTS

We consider the ensemble in Table I, and for each of the  $N_0$  configurations, we perform  $N_1 = 40$  Monte Carlo updates while keeping  $B$  fixed. The updates are separated by 60 sweeps, where one sweep is composed of eight over-relaxation updates followed by one heat-bath update. Both updates are performed using the Cabibbo-Marinari technique applied to three SU(2) subgroups [17,18].

In the following, we present our findings concerning the scaling of the errors with respect to  $N_1$  and show the application of our algorithm for the computation of the two-point correlation function over the whole range of distances allowed in our finite size lattice. The limitations of the method are also discussed. We conclude this section by using our method to compute the topological susceptibility and compare it to the result obtained with the standard algorithm.

### A. Autocorrelation times

An interesting question to explore is whether or not an undesirable growth of the autocorrelations is introduced due to the freezing of the boundary  $B$ . Such an effect could have an impact on the cost of the measurement in our nested Monte Carlo scheme. To investigate that, we look at the integrated autocorrelation time  $\tau_{\text{int}}$  of  $\mathcal{O}(x_0)$  as a function of the time coordinate  $x_0$ . Given that the  $N_0$  standard updates are completely decorrelated, the relevant autocorrelation function is given by the average over  $N_0$  of the autocorrelation function for the  $N_1$  nested updates,  $\tilde{\Gamma}(t) = \frac{1}{N_0} \sum_{i=1}^{N_0} \Gamma^i(t)$ , where  $\Gamma^i(t)$  is precisely the autocorrelation function for each of the nested chains.

Now,  $\tau_{\text{int}}$  can be defined in the usual way [16] in terms of the average autocorrelation function  $\tilde{\Gamma}(t)$ . Our data show that  $\tau_{\text{int}}$  increases at most by a factor of 1.5 when the observables approach the boundary  $B$ , so there is not a significant effect. However, on different observables, it could have a more severe impact which then must be taken into account when spacing the  $N_1$  nested updates and calculating the cost of the simulation.

### B. Choice of the parameters

The introduction of the nested updates adds an extra parameter to be tuned in the algorithm, as the parameter  $N_1$  can be chosen to minimize the errors at a given computational effort. We argue that for the connected correlator  $\hat{C}_O$ , when source and sink are placed far away from the boundary, the value of  $N_1$  up to which the algorithm is efficient can be scaled exponentially with respect to the distance to  $B$ .

To show this, in the Appendix, we have looked into the scaling of errors with respect to  $N_0$  and  $N_1$  in a Monte Carlo simulation. Our results show that the leading contribution to the error in the connected correlator scales as  $1/\sqrt{N_0 N_1}$ , which corresponds to the ideal case, but additionally there are other terms that scale as  $1/\sqrt{N_0 N_1}$  and as  $1/\sqrt{N_0}$ . Such terms, however, when dealing with connected correlation functions, are exponentially

suppressed as  $e^{-m_0|x_0^M - x_0^B|}$ , where  $x_0^M$  is the time coordinate of either the source or sink, whichever is the closest to the boundary, and  $m_0$  is the mass of the lightest mode which is compatible with the symmetries of  $\mathcal{O}$ . This means that we can expect the ideal scaling up to very large values of  $N_1$  given that source and sink are far away from the boundary in units of  $1/m_0$ .

Another effect that must be taken into account is the presence of the correction term  $\Delta$ . Such a term is measured only over the  $N_0$  standard updates, so that its error should scale in the standard way as  $1/\sqrt{N_0}$ . This will add another term which is independent of  $N_1$  to the final error. We can see from our results in Fig. 3 that for a fixed  $N_0$ , the error in  $\Delta$  decays at least exponentially fast with the distance of either the source or sink to the boundary  $B$ . This means that for the final estimator  $\hat{C}$ , the value of  $N_1$  up to which the ideal scaling is valid increases exponentially with the distance to the boundary as long as  $|x_0^M - x_0^B|$  is larger than the relevant scale, either  $1/m_0$  for the effects coming from  $\tilde{C}$  or  $\sqrt{8t}$  for those coming from  $\Delta$ .

### C. $N_1$ dependence of the error

To show the way in which our algorithm improves over the standard one, we measure the scaling of errors with respect to  $N_1$  for the symmetric correlator. The results for two different values of  $r$  are shown in Fig. 4. For the larger  $r = 28a$ , and for  $N_1 = 40$ , we are still in the regime where the ideal scaling is the dominant one, so on the left subplot, we see a scaling of the error which is compatible with  $1/N_1$  for the whole range of  $N_1$  values.

For the smaller value of  $r = 14a$ , in particular when looking at the case of  $\hat{C}_e$ , we observe that for  $N_1 \gtrsim 6$ , the error improves only marginally with  $N_1$ , which means that we are already in the regime where the term independent of  $N_1$  becomes relevant. This supports the discussion of the previous section and shows that for small values of  $r$ , there is no significant improvement by performing a very large number of  $N_1$  nested Monte Carlo updates. In practice, one can use all the  $N_1$  generated nested updates for all values of

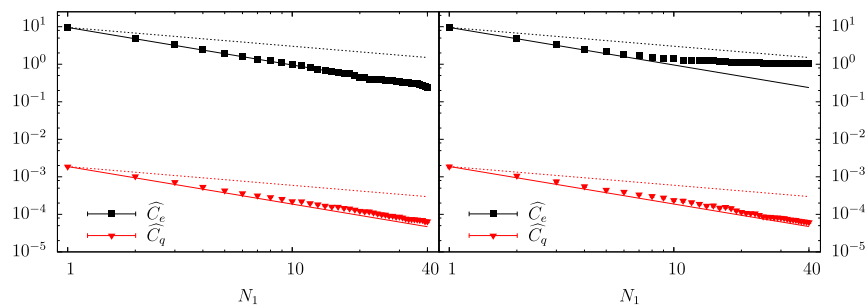


FIG. 4. Scaling of the error of  $\hat{C}_O$  as a function of  $N_1$ . On the left for a value of  $r = 28a = 13.2\sqrt{t_0}$  and on the right for a value of  $r = 14a = 6.6\sqrt{t_0}$ . The solid line indicates a scaling of the error proportional to  $1/N_1$ , while the dotted line corresponds to the standard  $1/\sqrt{N_1}$  scaling. For the smaller value of  $r$  (right plot), we observe a saturation in the number of effective  $N_1$  nested updates that can be used to reduce the errors. In fact, after  $N_1 \approx 6$  we observe no significant improvement.

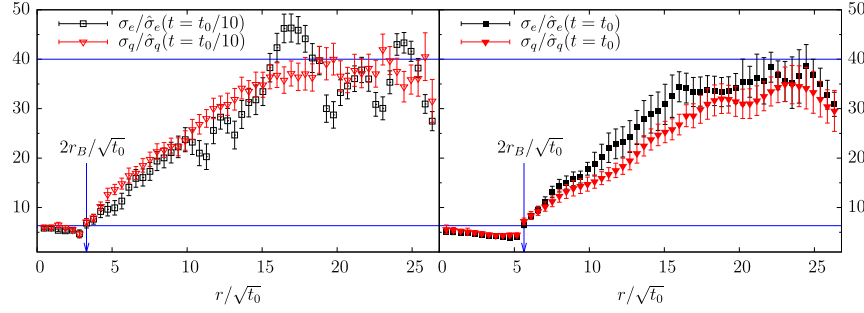


FIG. 5. Ratio of the errors  $\sigma_O/\hat{\sigma}_O$  as a function of  $r$ . Open symbols are the results for a flow time  $t = t_0/10$ , while filled symbols corresponds to the value of  $t = t_0$ . One can see that the improvement can be split into three distinct regions. For short distances, our algorithm is not as efficient as the standard one. For intermediate distances our algorithm is already better than the standard one but does not reach the theoretical maximum improvement, which is only achieved in the large distance regime. The two horizontal lines represent the theoretical maximum improvement of the standard algorithm and the one expected from our algorithm.

$r$ , but for small separations, the effect of using all of them is not significant.

For a given  $N_0$  and  $N_1$ , the value of  $r$  at which the ideal scaling is not valid anymore is observable dependent, and so it has to be studied on a case by case basis. In our particular case, we observe that for  $N_1 = 40$ , we are on the ideal scaling regime for the correlator at distances starting at values of  $r = 16a = 7.5\sqrt{t_0}$  at a flow time  $t = t_0$ .

#### D. Application of the algorithm

To show how the algorithm performs for the whole range of distances in the two-point correlator, we compute  $C_q$  and  $C_e$  using the standard algorithm and using our nested Monte Carlo scheme. For each value of  $r$  and  $x_0$ , we compute  $C_O$  and  $\hat{C}_O$ . We use the  $N_0 = 384$  standard updates to compute  $C_O$  in the usual way. For our nested algorithm, we employ the  $N_1 = 40$  nested updates for each of the standard ones.

When using the standard approach, the correlator at distance  $r$  is computed by averaging over all the  $x_0$  values in the plateau region. In the case of our algorithm, this is not the best strategy, as translation invariance is lost due to the presence of the boundary  $B$ . Instead, we find it beneficial not to use those time slices for which the source or sink is closer to  $B$  than a given distance  $r_B$ , which is tuned as part of the analysis. When working at  $t = t_0$ , we find the best choice to be  $r_B = 6a$ , which is compatible with the smearing radius  $\sqrt{8t_0} \approx 6a$ .

The inclusion of  $r_B$  in the analysis means that for separations smaller than  $2r_B$ , the average is done only when the source and sink are in the same domain, either  $L$  or  $R$ . In those cases, we expect no improvement with respect to the standard algorithm. For larger distances, however, one can choose to have  $x_0 \in L$  and  $x_0 + r \in R$ , where the better scaling is expected. Notice that for intermediate distances, the average over time slices would also include terms for which the source and sink are in the same domain. These terms would contribute to the error

with the usual scaling  $1/\sqrt{N_0N_1}$ , so we find the better performance when they are also not included in the average, and we sum only over the factorized terms.

We also look at smaller values of the flow time  $t$ ; in particular, we look at a value of  $t = t_0/10$ . Smaller flow times can be of interest if one is looking at obtaining the glueball masses. In such cases, the analysis is the same as described above, but only the value of  $r_B$  changes; for example, at  $t = t_0/10$ , we find an optimal value of  $r_B = 3a$ , which is also compatible with the value of the smearing radius.

#### E. Performance of the algorithm

We apply the strategy described above to compute the  $\hat{C}_O$  and  $C_O$  correlators for a wide range of separations  $r$  between the source and sink. To assess the performance of the algorithm, in Fig. 5 we plot the ratio between the error of the standard correlator  $\sigma_O$  and the error of the improved one  $\hat{\sigma}_O = \sigma(\hat{C}_O)$ . With the standard algorithm, if the statistics are increased by a factor  $N_1 = 40$ , the error should scale down by a factor  $\sqrt{N_1} \approx 6.3$ . The lower horizontal line in Fig. 5 shows the theoretical improvement of the standard algorithm for the same statistics as the ones we use in our two-level algorithm.

For the short distance region, we observe an improvement which is below the theoretical one of the standard algorithm. As explained before, this is expected due to the fact that one cannot make full use of translation invariance and that our algorithm is not designed to be the most efficient for such short distances when the effects of the flow are more relevant.

As soon as  $r \geq 2r_B$ , one enters the region where the new algorithm outperforms the standard one. This is expected, as for most of these values of  $r$ , we can make full use of the  $N_1 = 40$  nested updates. However, at intermediate distances, we lose due to the lack of translation invariance in the  $x_0$  direction. This is precisely what we observe as the improvement rises continually from  $r = 2r_B$  until it reaches

the theoretical maximum improvement equal to  $N_1 = 40$ . For values of  $r$  sufficiently large, our algorithm performs as expected, and we obtain the theoretical maximum improvement which is shown in the figure by the upper horizontal line. We observe the same qualitative behavior for different values of the flow time, the only difference being the different value of  $r_B$  which is used in the analysis. Clearly, for smaller values of the flow time, we are able to outperform the standard algorithm at even shorter distances, which could be useful for certain applications.

### F. Topological susceptibility

As a final test of our proposal, we compute the topological susceptibility  $\chi$  at  $t = t_0$  and compare it to the result obtained when using the standard algorithm. For the comparison, we use the same statistics in both cases, i.e.,  $N_0 N_1 = 15360$  measurements, so that the computational effort is roughly the same. For the definition of the susceptibility, we use the one in Ref. [7]. To write this in terms of our observables, we define  $\bar{C}_q(r)$  as the average of  $C_q(x_0, r)$  over  $x_0$ . We proceed as described in the previous section, so for the standard algorithm, we average over all values of  $x_0$ , while in the case of the new algorithm, we use only those values of  $x_0$  such that the source and sink are not closer than  $r_B$  to the boundary  $B$ . Then, we define the topological susceptibility as

$$\chi(r_{\text{cut}}) = \frac{a}{L^3} \sum_{z_0=-r_{\text{cut}}}^{r_{\text{cut}}} \bar{C}_q(|z_0|), \quad (9)$$

where  $r_{\text{cut}}$  should be chosen so that the statistical error in the sum is larger than the estimated systematic error from cutting the summation. We are not so interested in choosing the best value of  $r_{\text{cut}}$  but more on comparing the performance of the two-level algorithm with respect to the standard one.

In Table II, we show the results at three different values of  $r_{\text{cut}}$  using both the standard algorithm and the new nested Monte Carlo algorithm that we propose in this paper. As already pointed out in the Introduction, with the traditional approach, summing up the correlator to large values of  $r_{\text{cut}}$  only increases the error while the signal remains relatively constant [6,7]. We clearly observe this effect in our data when using the standard method. On the other hand, the

TABLE II. Results for the topological susceptibility  $10^4 t_0^2 \chi(r_{\text{cut}})$  using the standard algorithm and the new algorithm that we propose in this paper. The values of  $r_{\text{cut}}$  in physical units were computed using the  $r_0$  scale from Ref. [15].

$r_{\text{cut}}/\sqrt{t_0}$	$r_{\text{cut}}$ (fm)	Standard	New
5.1	0.85	6.405(46)	6.347(60)
15.4	2.56	6.507(94)	6.291(61)
25.2	4.19	6.518(164)	6.254(69)

error when using our algorithm remains relatively constant when the value of  $r_{\text{cut}}$  is increased from values of 0.85 fm up to 4.19 fm. In fact, for the largest value of  $r_{\text{cut}}$ , the improvement when using our algorithm is more than twofold, corresponding to an increase in statistics by a factor 5.

## V. CONCLUSION

In this paper, we have studied a multilevel algorithm for computing the two-point correlation function of flow observables. It is based on the idea originally introduced in Ref. [10]. Basically, we split the lattice into two subvolumes separated by a boundary  $B$  and use the locality of the action to perform independent updates on each of them. Such an approach would not work for observables at positive flow time, so we slightly modify the flow equations to build a “good” approximation of the original observable which can be factorized as required for a multilevel type scheme to work.

In this type of algorithms, one starts by performing  $N_0$  standard updates followed by  $N_1$  nested updates for each of the original  $N_0$  generated configurations. In the ideal case, one expects the scaling of the error to be proportional to  $1/N_1$  instead of the standard  $1/\sqrt{N_1}$ . We put this to the test, and for the case of the connected two-point correlation function  $\langle \mathcal{O}(x)\mathcal{O}(y) \rangle - \langle \mathcal{O}(x) \rangle \langle \mathcal{O}(y) \rangle$ , we find that our algorithm outperforms the standard one when  $x$  and  $y$  are far from the boundary  $B$  in units of  $1/m_0$  and of the flow radius  $\sqrt{8t}$ , where  $m_0$  is the lightest mass compatible with the observable  $\mathcal{O}$ . In the case of short separations, our algorithm is not better than the standard one, which is expected from the way the observables are constructed.

We also showed that our algorithm can be used to obtain a better lattice determination of the topological susceptibility  $\chi$ , where the large statistical errors coming from the tail of the correlator are tamed. With our choice of parameters, we observe a decrease of errors by a factor larger than 2 for the same statistics as the standard algorithm, which would correspond to a fivefold decrease of the computational time required for a fixed target error.

Although we performed our analysis with the Yang-Mills energy density  $e$  and the topological charge  $q$ , the idea can be applied to any correlation function of flow observables in the lattice Yang-Mills gauge theory. Also, the idea that we presented in this paper can be generalized to a four dimensional approach in which the decomposition is not limited to the time direction. In that case, we expect an even better performance of the algorithm.

## ACKNOWLEDGMENTS

We are very thankful to R. Sommer for extensive discussions. We also would like to thank L. Giusti, M. Cè, and D. Banerjee for discussions related to multilevel algorithms. Our simulations were performed at the ZIB

computer center with the computer resources granted by The North-German Supercomputing Alliance. M. G. V. acknowledges support from the Research Training Group GRK1504/2 “Mass, Spectrum, Symmetry” founded by the German Research Foundation.

### APPENDIX: ERROR REDUCTION

In a two-level nested Monte Carlo algorithm as the one described in the main text, we are interested in the scaling of errors with respect to  $N_0$  and  $N_1$ . In particular, we look at the case of the two-point correlator

$$A = \langle \mathcal{O}(x_0)\mathcal{O}(y_0) \rangle,$$

where  $\mathcal{O}(x_0) \in L$  and  $\mathcal{O}(y_0) \in R$ . To simplify the notation, we write  $\mathcal{O} \equiv \mathcal{O}(x_0)$  and  $\mathcal{O}' \equiv \mathcal{O}(y_0)$ .

In a Monte Carlo simulation, an estimator for  $A$  is given by

$$\hat{A} = \frac{1}{N_0} \sum_{i=1}^{N_0} \frac{1}{N_1^2} \sum_{j=1}^{N_1} \sum_{k=1}^{N_1} \mathcal{O}^{ij} \mathcal{O}'^{ik}. \quad (\text{A1})$$

The error  $\sigma_A^2$  on the estimator is then computed in the usual way,

$$\sigma_A^2 = \langle (\hat{A} - \bar{A})^2 \rangle_{LBR}, \quad (\text{A2})$$

where  $\langle \rangle_{LBR}$  stands for the average over all the gauge links in  $L \cup B \cup R$  and  $\bar{A} = \langle [\mathcal{O}]_L [\mathcal{O}' ]_R \rangle_B$  is the real expectation value of  $A$ .

By inserting  $\hat{A}$  from Eq. (A1) into Eq. (A2) and using the fact that the  $N_0$  updates are independent, one obtains

$$\sigma_A^2 = \frac{1}{N_0 N_1^2} \langle \text{Var}_L(\mathcal{O}) \text{Var}_R(\mathcal{O}') \rangle_B + \frac{1}{N_0} (\langle [\mathcal{O}]_L^2 [\mathcal{O}' ]_R^2 \rangle_B - \bar{A}^2) + \frac{1}{N_0 N_1} (\langle \text{Var}_L(\mathcal{O}) [\mathcal{O}' ]_R^2 + \text{Var}_R(\mathcal{O}') [\mathcal{O}]_L^2 \rangle_B), \quad (\text{A3})$$

where  $\text{Var}_L(\mathcal{O}) = [\mathcal{O}^2]_L - [\mathcal{O}]_L^2$  and similarly for  $\text{Var}_R(\mathcal{O}')$ . By looking at Eq. (A3), it is clear that the error scales not only as the ideal case  $1/\sqrt{N_0 N_1}$ , but it has also subleading contributions.

Note, however, that using the transfer matrix formalism, one can show that the second term proportional to  $1/N_0$  is exponentially suppressed as  $e^{-m_0|x_0^B - x_0^M|}$ , where  $m_0$  is the mass of the lightest state compatible with the symmetries of  $\mathcal{O}$  and  $x_0^M$  corresponds to  $x_0$  or  $y_0$ , whichever is the closest to  $x_0^B$ .

The third term is also exponentially suppressed if one considers the case of the connected correlator

$$C = \langle \mathcal{O}(x_0)\mathcal{O}(y_0) \rangle - \langle \mathcal{O}(x_0) \rangle \langle \mathcal{O}(y_0) \rangle.$$

Then, only the first term gives the leading contribution to the error, and it is the one that has the ideal scaling for which a nested Monte Carlo scheme would be useful.

The final formula for the error of the connected correlator is

$$\sigma_C^2 \approx \frac{1}{N_0 N_1^2} \langle \text{Var}_L(\mathcal{O}) \text{Var}_R(\mathcal{O}') \rangle_B + e^{-m_0|x_0^B - x_0^M|} \left( \frac{c_1}{N_0 N_1} + \frac{c_2}{N_0} \right). \quad (\text{A4})$$

- 
- [1] G. Parisi, The strategy for computing the hadronic mass spectrum, *Phys. Rep.* **103**, 203 (1984).
- [2] R. Narayanan and H. Neuberger, Infinite N phase transitions in continuum Wilson loop operators, *J. High Energy Phys.* **03** (2006) 064.
- [3] M. Lüscher, Properties and uses of the Wilson flow in lattice QCD, *J. High Energy Phys.* **08** (2010) 071.
- [4] M. Lüscher and P. Weisz, Perturbative analysis of the gradient flow in non-abelian gauge theories, *J. High Energy Phys.* **02** (2011) 051.
- [5] M. Cè, C. Consonni, G. P. Engel, and L. Giusti, Non-Gaussianities in the topological charge distribution of the SU(3) Yang–Mills theory, *Phys. Rev. D* **92**, 074502 (2015).
- [6] A. Bazavov *et al.* (MILC Collaboration), Topological susceptibility with the asqtad action, *Phys. Rev. D* **81**, 114501 (2010).
- [7] M. Bruno, S. Schaefer, and R. Sommer (ALPHA Collaboration), Topological susceptibility and the sampling of field space in  $N_f = 2$  lattice QCD simulations, *J. High Energy Phys.* **08** (2014) 150.
- [8] A. Chowdhury, A. Harindranath, and J. Maiti, Open boundary condition, Wilson flow and the scalar glueball mass, *J. High Energy Phys.* **06** (2014) 067.
- [9] G. Parisi, R. Petronzio, and F. Rapuano, A measurement of the string tension near the continuum limit, *Phys. Lett.* **128B**, 418 (1983).
- [10] M. Lüscher and P. Weisz, Locality and exponential error reduction in numerical lattice gauge theory, *J. High Energy Phys.* **09** (2001) 010.
- [11] M. Cè, L. Giusti, and S. Schaefer, Domain decomposition, multi-level integration and exponential noise reduction in lattice QCD, [arXiv:1601.04587](https://arxiv.org/abs/1601.04587).



- [12] H. B. Meyer, Locality and statistical error reduction on correlation functions, *J. High Energy Phys.* **01** (2003) 048.
- [13] M. Lüscher, Trivializing maps, the Wilson flow and the HMC algorithm, *Commun. Math. Phys.* **293**, 899 (2010).
- [14] M. Lüscher and S. Schaefer, Lattice QCD without topology barriers, *J. High Energy Phys.* **07** (2011) 036.
- [15] S. Necco and R. Sommer, The  $N(f) = 0$  heavy quark potential from short to intermediate distances, *Nucl. Phys.* **B622**, 328 (2002).
- [16] U. Wolff (ALPHA Collaboration), Monte Carlo errors with less errors, *Comput. Phys. Commun.* **156**, 143 (2004); **176**, 383(E) (2007).
- [17] N. Cabibbo and E. Marinari, A new method for updating SU (N) matrices in computer simulations of gauge theories, *Phys. Lett.* **119B**, 387 (1982).
- [18] F. R. Brown and T. J. Woch, Overrelaxed Heat Bath and Metropolis Algorithms for Accelerating Pure Gauge Monte Carlo Calculations, *Phys. Rev. Lett.* **58**, 2394 (1987).